arXiv:1203.3984v2 [math.PR] 6 Sep 2012

On the geometric ergodicity of nonlinear multivariate time series

Marco Ferrante *
Dipartimento di Matematica
Università degli Studi di Padova
via Trieste 63
35121 Padova, Italy
e-mail: ferrante@math.unipd.it

Giovanni Fonseca
Dip. di Scienze Economiche e Stat.
Università degli Studi di Udine
via Tomadini, 30/A
33100 Udine, Italy
e-mail: giovanni.fonseca@uniud.it

Abstract

In this paper we consider multivariate time series obtained as solution to multidimensional nonlinear stochastic difference equations, whose coefficients are allowed to be locally degenerate and to present discontinuities. We provide simple and easy to check sufficient conditions for the irreducibility, T-chain regularity and geometric ergodicity of these processes and apply the results to the BEKK-ARCH(1) models with a nonlinear autoregressive term.

Keywords: Nonlinear stochastic difference equation, T-chain, Geometric ergodicity, BEKK-ARCH(1).

AMS Classification: 60G10, 60J05, 62M10

Short title: Ergodicity of multivariate models

1 Introduction

Let us consider a system of nonlinear stochastic difference equations

$$X_t = f(X_{t-1}) + g(X_{t-1})e_t, \quad t \geq 1$$

*corresponding author

where $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^{n\times k}$, $\{e_t, t \in \mathbb{N}\}$ is a sequence of independent, identically distributed k-dimensional random vectors and X_0 is a given random vector. We would like to find simple and easy to check conditions that ensure the solution to be an irreducible, T-chain and a geometric ergodic process. The interest on these kind of models is clear: first of all they can be thought as the discretization of a multidimensional stochastic differential equation and the properties of the discretized models are usually of extreme interest. Moreover, from a statistical point of view, they can be considered as a state space representation of a first order multivariate time series model and many examples in the literature can be represented by these models.

At a first sight this problem does not look very original and worth to be studied in a new paper. However, under some very natural conditions on the coefficients f and g, like absence of continuity and possible local singularity of the matrix valued function g, to the best of our knowledge no general results have been published so far, except for those present in our former paper [5], in the case $n = k = 1$. It is worth to remark that several paper deal with similar or more general models (see e.g. [1], [2], [9], [10] and [12]), but in all these papers stronger assumptions are required on g, as everywhere continuity and non degeneracy.

The aim of this paper is to provide a first step in order to fill this gap in the literature. We will restrict ourselves to the case $n = k$ and we will assume that the noise random variables e_t possess a strictly positive density on $\mathbb{R}^n$. Under these conditions, we will be able to obtain for this class of models results similar to those that hold in the regular case, adapting some of the standard techniques applied to the smooth version of the present equation. As an application, we shall consider a BEKK-ARCH(1) multivariate model (see Engle and Kroner [4], Hansen and Rahbek [7] and Saikkonon [12]) and obtain a set of sufficient condition to be this process geometric ergodic.

The paper is organized as follows: in Section 2 we will present the model and recall some notation and known results. In Section 3 we will consider the problem to find out sufficient conditions to be the solution an aperiodic, irreducible, T-chain. This part is fairly technical, but these three properties are the fundamental ingredient in order to apply the well known

Foster-Lyapounov drift criteria of Section 4. This technique allows us to determine a set of sufficient conditions to be the solution an ergodic process. In the last section we shall apply the results to a BEKK-ARCH(1) model with a general nonlinear autoregressive term.

By λ_n we will denote the Lebesgue measure on $\mathbb{R}^n$. By A^o we will denote the interior of the set A. For $p \geq 1$, $\|\cdot\|_p$ will denote the l_p norm on $\mathbb{R}^n$. For $0 < s \leq 1$ and $x \in \mathbb{R}^n$, we will define $\|x\|_s = \sum_{i=1}^n |x_i|^s$; this is clearly no more a norm, but it still defines a pseudometric on $\mathbb{R}^n$, since the triangular inequality holds true. With $|||\cdot|||$ we will denote a generic matrix norm on $\mathbb{R}^{n\times n}$; for $p \geq 1$, $|||\cdot|||_p$ will denote the operator norm associated with $\|\cdot\|_p$, while $|||\cdot|||_{1,p}$ will denote the *maximum column sum* matrix norm associated with $\|\cdot\|_p$, whose definition is as follows:

$$|||A|||_{1,p} = \max_{1\leq j\leq n} \|a_{\cdot j}\|_p \quad .$$

We will use the same definition when $0 < s \leq 1$, even if again the function $|||\cdot|||_{1,s}$ will not be a norm anymore. Finally, by $|||\cdot|||_F$ we will denote the Frobenius norm

$$|||A|||_F = \left(\sum_{1\leq i,j\leq n} |a_{ij}|^2 \right)^{1/2}$$

(see [8], Section 5.6 for a complete account on this topic).

For a map $G : X \times Y \to Z$, we shall denote by G_x the x-section of G, namely $G_x(y) := G(x,y)$, while, given $B_1, \dots, B_t \in \mathbb{R}^n$, we shall denote $B_{1:t} = B_1 \times \cdots \times B_t$ and, similarly, $u_{1:t} = (u_1, u_2, \dots, u_t) \in \mathbb{R}^{nt}$.

2 The multidimensional stochastic difference equation

In this paper we will study nonlinear stochastic difference equations defined by the system

$$\begin{cases} X_t = f(X_{t-1}) + g(X_{t-1})e_t, \quad t \geq 1 \\ X_0 = \xi \end{cases} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, $\{e_t, t \in \mathbb{N}\}$ is a sequence of independent, identically distributed n-dimensional random vectors and ξ is a given random vector. In the discrete time case we have clearly no problems related to the existence and uniqueness of the solution, while a fundamental question is if the system is ergodic, which is related to the fact that it admits an invariant distribution. We will be able to prove that, under some assumptions on the coefficients f and g and on the law of the noise sequence, the solution to (1) is geometrically ergodic.

Let us start by considering the regularity of the coefficients f and g. In this paper, we consider the case with both the coefficients not everywhere continuous and the matrix $g(x)$ locally singular. This last assumption is not weird: in the scalar case this means that the function g could be zero somewhere and this is indeed the case if we choose g to be an affine function. Nevertheless in the literature g is usually assumed to be non singular and a possible set of hypothesis (see Liebscher [10]) is that there exist two constants $C_1, C_2 > 0$ such that $|||g^{-1}(x)||| \leq C_1$ and $|\det(g(x))| \leq C_2$ on every compact subset of $\mathbb{R}^n$ (see also Saikkonon [12] for the BEKK-ARCH model).

In the present paper, denoting by $\Theta := \{x \in \mathbb{R}^n : \det(g(x)) \neq 0\}$ the set of "regular" points of g and by $\mathcal{C}_f$ (resp. $\mathcal{C}_g$) the set of the continuity points of the function f (resp. g), we will require that the following two assumptions are satisfied:

(H.1) f and g are locally bounded and the sets Θ, $\mathcal{C}_f$ and $\mathcal{C}_g$ have not empty interior.

Under smooth conditions on the coefficients f and g, one fruitful approach is to use the concept of the forward accessibility from the control theory and its equivalence with the much more workable Rank condition (see e.g. Meyn-Tweedie [11], Chapter 7). Since in our case we do not assume differentiability of the coefficients, we have to find out a different approach, even if our property will be at the end stronger, but not so far from the forward accessibility. Let us denote by $O \subseteq \mathbb{R}^n$ the support of the random vector e_1, $F(x, u) = f(x) + g(x)u$ and, inductively, for $t \in \mathbb{N}^+$

$$F^{t+1}(x_0, u_1, \ldots, u_{t+1}) := F(F^t(x_0, u_1, \ldots, u_t), u_{t+1}) \quad . \tag{2}$$

When $t = 0$, $F^t(x_0, u_1, \ldots, u_t) \equiv x_0$.

Our second assumption will be that

(H.2) For any $x_0 \notin \Theta$, there exists $t \in \mathbb{N}^+$ and $u_1, \ldots, u_t \in O$ such that $F^t(x_0, u_1, \ldots, u_t) \in (\Theta \cap \mathcal{C}_f \cap \mathcal{C}_g)^o$ and $F^{t-1}(x_0, u_1, \ldots, u_{t-1}) \in \mathcal{C}_f \cap \mathcal{C}_g$.

Remark 1 *Under (H.2) and assuming that O has non-empty interior, we easily obtain that for any x_0, there exists $t \in \mathbb{N}^+$ and $u_1, \ldots, u_t \in O$ such that $F^{t+1}_{x_0,u_1,\ldots,u_t}(O)$ has non-empty interior. This condition implies, but is evidently stronger than the* forward accessibility*, which requires that for any $x_0 \in \mathbb{R}^n$, $\bigcup_{t=0}^{+\infty} F^t_{x_0}(O^t)$ has non-empty interior.*

Remark 2 *If we assume that f and g are continuous, a sufficient condition in order to satisfy assumption (H.2) is that for any $x \in \mathbb{R}^n$, there exists $t \in \mathbb{N}$ such that $f^t(x) \in \Theta^o$.*

To conclude, let us state the assumptions on the noise sequence $\{e_t, t \in \mathbb{N}\}$. The price to be paid for the weak hypothesis (H.1) is quite expensive, since we have to assume absolute continuity and lower semicontinuity of the noise density. However, this is often the additional condition that we have to ask in order to allow some kind of singularity in the coefficient g (see e.g the results on the bilinear processes in [13] and [3], the threshold bilinear processes in [5] and the nonlinear state space models in [11]).

(H.3) $\{e_t, t \in \mathbb{N}\}$ is a sequence of independent, identically distributed n-dimensional random vector, absolutely continuous w.r.t. Lebesgue measure λ_n on $\mathcal{B}(\mathbb{R}^n)$, with density $p(\cdot)$ strictly positive almost everywhere and lower semicontinuous.

3 Irreducibility, aperiodicity and T-chain property

In order to apply the classic Foster-Lyapunov drift criteria of the next section (see Meyn-Tweedie [11] for a comprehensive introduction to this topic), we need three basic ingredients. Indeed, we have to prove that the Markov chain, solution to (1), is φ-irreducible, for a given measure φ, aperiodic and a T-chain. While it is usually not too difficult to find out a set of reasonable conditions that ensure that the process is φ-irreducible and aperiodic, it is more challenging to handle the T-chain condition. Most of the papers in the literature do not spend much time on

this part of the study and the authors usually state some general conditions that ensure the process to satisfy these three properties (see e.g. Liebscher [10], Section 4). Since in this paper we would like to allow the diffusion coefficient g to be locally singular, we shall need to impose the previous set of stronger assumptions (H.1)-(H.3).

Let us start by stating a simple result, that we will need in the sequel and whose proof is immediate (see [5]).

Lemma 3 *Let $A, B \subseteq \mathbb{R}^n$: if $F : A \to B$ is a continuous function and $G, H : B \to \mathbb{R}$ are two lower semicontinuous (lsc) functions, then $G \circ F$ and $G \cdot H$ are lsc.*

We are now able to prove the main result of this section:

Proposition 4 *Under $(H.1) - (H.3)$, the process solution to (1) is a λ_n-irreducible, aperiodic T-chain.*

Proof. **λ_n-irreducibility**: we have to prove that for any $A \in \mathcal{B}(\mathbb{R}^n)$, such that $\lambda_n(A) > 0$, and any $x \in \mathbb{R}^n$, there exists $t = t(x, A) \in \mathbb{N}$ such that $P^t(x, A) = \mathbb{P}[X_t \in A | X_0 = x] > 0$. If $\det(g(x)) \neq 0$, we get $P(x, A) > 0$. Otherwise, by assumption (H.2) we get that there exists $t \in \mathbb{N}$ and $u_1, \ldots, u_t \in \mathbb{R}^n$ such that $F^t(x, u_1, \ldots, u_t)$, defined in (2), belongs to $(\Theta \cap \mathcal{C}_g)^o$, and is continuous in $(x, u_1, \ldots, u_t)$. Therefore, there exist open balls $B_1, \ldots, B_t$ in $\mathbb{R}^n$ such that $F^t(x, B_1, \ldots, B_t) \subseteq \Theta^o$. By (H.1)-(H.3) we get

$$P^{t+1}(x, A) = \mathbb{P}\left[f(F^t(x, e_{1:t})) + g(F^t(x, e_{1:t}))e_{t+1} \in A\right]$$

$$\geq \int_{B_{1:t}} \Big[\int_A \left|\det(g(F^t(x, u_{1:t})))\right|^{-1} p(g(F^t(x, u_{1:t}))^{-1}(u_{t+1} - f(F^t(x, u_{1:t}))))du_{t+1}\Big] \times$$

$$p(u_1) \cdots p(u_t) du_{1:t} \geq c_3 \int_A \Big[\int_{B_{1:t}} p(g(F^t(x, u_{1:t}))^{-1}(u_{t+1} - f(F^t(x, u_{1:t})))) \times$$

$$p(u_1) \cdots p(u_t) du_{1:t}\Big] du_{t+1} > 0 \quad ,$$

where $c_3 := \inf_{u_{1:t} \in B_{1:t}} \left|\det(g(F^t(x, u_{1:t})))\right|^{-1} < \infty$, and the λ_n-irreducibility is proved.

Aperiodicity: we will prove that the solution process is strongly aperiodic, i.e. that there exist a nontrivial measure ν_1 on $\mathcal{B}(\mathbb{R}^n)$ and a subset $A \in \mathcal{B}(\mathbb{R}^n)$, with $\nu_1(A) > 0$, such that for any $x \in A$ and $B \in \mathcal{B}(\mathbb{R}^n)$, $P(x, B) \geq \nu_1(B)$. Let us take $x \in (\Theta \cap \mathcal{C}_f \cap \mathcal{C}_g)^o$; by the assumption (H.1) we get that there exists an open bounded neighborhood A of x and two positive constants c_1, c_2 such that

$$0 < c_1 \leq |\det(g(y))| \leq c_2$$

for any $y \in A$. By (H.2) we get

$$P(x, B) = \int_B |\det(g(x))|^{-1} p(g(x)^{-1}(y - f(x)))dy$$

$$\geq c_2^{-1} \int_{A \cap B} p(g(x)^{-1}(y - f(x)))dy \geq c_2^{-1} k_1 \lambda_n(A \cap B) \quad ,$$

where $0 < k_1 = \inf\limits_{x,y \in A} p(g(x)^{-1}(y - f(x)))$. The result holds for $\nu_1(\cdot) = c_2^{-1} k_1 \lambda_n(A \cap \cdot)$.

T-chain condition: By Proposition 6.4.2 in Meyn and Tweedie [11], it will be sufficient to prove that for each $x \in \mathbb{R}^n$, there exists a $t \in \mathbb{N}$ and a non trivial substochastic transition kernel $T_x(\cdot, \cdot)$. l.s.c. in the first variable, such that $P^t(y, A) \geq T_x(y, A)$ for each $y \in \mathbb{R}^n$ and $A \in \mathcal{B}(\mathbb{R}^n)$. Let $x \in \mathbb{R}^n$: by assumption (H.2) we get that there exist $t \in \mathbb{N}$ and $u_1, \ldots, u_t \in \mathbb{R}^n$ such that $F^t(x, u_1, \ldots, u_t) \in (\Theta \cap \mathcal{C}_f \cap \mathcal{C}_g)^o$ and $t + 1$ open bounded sets $B_0, B_1, \ldots, B_t$ of $\mathbb{R}^n$, such that $F^t(B_0, B_1, \ldots, B_t) \subseteq \Theta^o$. Moreover, we can assume that f and g are continuous on $F^t(B_0, B_1, \ldots, B_t)$. Hence, for $y \in B_0$ and $A \in \mathcal{B}(\mathbb{R}^n)$, we get

$$P^{t+1}(y, A) = \mathbb{P}\left[f(F^t(y, e_{1:t})) + g(F^t(y, e_{1:t}))e_{t+1} \in A\right]$$

$$\geq \int_{B_{1:t}} \left[\int_A \left|\det(g(F^t(y, u_{1:t})))\right|^{-1} p(g(F^t(y, u_{1:t}))^{-1}(u_{t+1} - f(F^t(y, u_{1:t}))))du_{t+1}\right] \times$$

$$p(u_1) \cdots p(u_t)du_{1:t} \geq c_4 \int_A \left[\int_{B_{1:t}} p(g(F^t(y, u_{1:t}))^{-1}(u_{t+1} - f(F^t(y, u_{1:t})))) \times\right.$$

$$\left. p(u_1) \cdots p(u_t)du_{1:t}\right] du_{t+1} =: \widetilde{T}(y, A) \quad ,$$

where $c_4 := \inf_{(y,u_{1:t})\in B_{0:t}} \left|\det(g(F^t(y,u_{1:t})))\right|^{-1} < \infty$. By Lemma 3 and Fatou's Lemma we get that $\widetilde{T}(y,A)$ (for $y \in B_0$) is a lsc function and we can define the substochastic transition kernel $T_x(y,A) := \phi(y)\widetilde{T}(y,A)$, with $\phi(\cdot)$ a smooth function whose support is contained in B_0. It is clear that $P^t(y,A) \geq T_x(y,A)$ for each $y \in \mathbb{R}^n$ and $A \in \mathcal{B}(\mathbb{R}^n)$ and the proof is complete. ∎

We conclude this section by considering a simple bivariate time-series, solution of a two dimensional difference equation, where is present a threshold and a singular part.

Example 5 *Let us take $n = 2$ and consider the difference equation (1), with f and g defined as follows*

$$f(x,y) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$g(x,y) = \begin{pmatrix} d_{11}\cdot x & d_{12}\cdot y \\ d_{21}\cdot x & d_{22}\cdot y \end{pmatrix} \mathbf{1}_{\mathbb{R}^2\setminus C} + \begin{pmatrix} d_{31}\cdot x & 0 \\ d_{32}\cdot y & 0 \end{pmatrix} \mathbf{1}_C + \begin{pmatrix} d_{41} & 0 \\ d_{42} & 0 \end{pmatrix}$$

where $C = \{x \leq 0, y \leq 0\}$ and $d_{11}d_{22} - d_{12}d_{21} \neq 0$. Note that the matrix $g(x,y)$ is singular for $(x,y) \in C \cup D_1$ and could be singular for $(x,y) \in D_2$, where $D_1 := \{(x,y) : x = 0, y > 0\}$ and $D_2 := \{(x,y) : x > 0, y = 0\}$. Moreover, g is not continuous on the boundary of C. It is easy to prove that if f and g are such that

$$d_{11}d_{42} - d_{21}d_{41} \neq 0 \ , \ d_{31}d_{42} - d_{32}d_{41} \neq 0 \tag{3}$$

and the noise sequence admits $\mathbb{R}^n$ as its support, then the hypotheses (H.1)-(H.2) are satisfied. Indeed, the only non trivial part is (H.2), which is always satisfied since, for $(x,y) \in C$, there exists $u \in \mathbb{R}^2$ such that $f(x,y) + g(x,y)u \notin C \cup D_1 \cup D_2$, while for $(x,y) \in D_i$, $i = 1,2$, there exist $u,v \in \mathbb{R}^2$ such that $f(f(x,y) + g(x,y)u) + g((f(x,y) + g(x,y)u))v \notin D_i$.

4 Geometric ergodicity

In this section we will obtain, in a standard way, a set of sufficient conditions on the coefficients of the difference equation in order to be the solution process geometrically ergodic. Due to the weak assumptions on this coefficients, we will obtain a rather strict sufficient condition, but this is in line with previous results in the literature. Since our approach to prove the geometric ergodicity is based on the choice of a drift function, for specific models like the BEKK-ARCH(1) model in the next section, it could be more convenient to use a different function, but the proof will be similar to the one presented here.

Let us consider the process $\{X_t, t \geq 0\}$ solution to equation (1) and assume that it is a λ_n-irreducible, aperiodic T-chain. In order to apply the classic Foster-Lyapunov drift criteria for $V(x) = 1 + \|x\|_s$, when $s > 0$, we will need to apply some easy properties of the functions $\|\cdot\|_s$ and $|||\cdot|||_{1,s}$. Given A, B two $n \times n$ real matrices and $x \in \mathbb{R}^n$, it holds that

$$|||AB|||_{1,s} \leq |||A|||_{1,s} \, |||B|||_{1,s}$$

and that

$$\|Ax\|_s \leq |||A|||_{1,s} \, \|x\|_s \quad ,$$

for any $s > 0$; these results are well known for $s \geq 1$ and immediate to be proven for $s < 1$. Note that, the same properties hold true for $|||\cdot|||_s$ instead of $|||\cdot|||_{1,s}$, when $s \geq 1$, and furthermore, when $s = 2$, for $|||\cdot|||_F$ instead of $|||\cdot|||_{1,2}$.

We are now ready to provide an easy to check set of sufficient conditions in order to be the solution process geometric ergodic.

Proposition 6 *Let $\{X_t\}$ be the solution process of (1) and assume that it is λ_n-irreducible, aperiodic, T-chain. If for $s > 0$*

i. f and g are locally bounded;

ii. there exist $a_f \geq 0$ and $M, a_g, b_f, b_g > 0$ such that

$$\|f(x)\|_s \leq a_f + b_f\|x\|_s \quad , \quad \|\!|g(x)|\!\|_{1,s} \leq a_g + b_g\|x\|_s$$

for any $x \in \mathbb{R}^n$ with $\|x\|_s > M$;

iii. $\gamma = b_f \; + \; b_g \; \mathbb{E}\left[\|e_1\|_s\right] < 1$;

then $\{X_t, t \in \mathbb{N}\}$ is geometrically ergodic.

Furthermore, if the previous conditions hold for $s \geq 1$, then each component of the stationary distribution has finite moments up to order s.

Proof. Let us consider the function $V(x) = 1 + \|x\|_s$ for an arbitrary $s > 0$ and the compact set $C = \{x \in \mathbb{R}^n : \|x\|_s \leq M\}$. Since the solution is a λ_n-irreducible T-chain, we have that C is also *petite.*

By triangular inequality and assumptions i.–iii. we obtain

$$\begin{aligned}\mathbb{E}[V(X_t)|X_{t-1} = x] &\leq (b_f + b_g\mathbb{E}[\|e_t\|_s])(1 + \|x\|_s) + \\ &+ a_f + a_g\mathbb{E}[\|e_t\|_s] + 1 - (b_f + b_g\mathbb{E}[\|e_t\|_s])\end{aligned}$$

for every $x \in \mathbb{R}^n$ such that $\|x\|_s > M$, and

$$\mathbb{E}[V(X_t)|X_{t-1} = x] \leq b_M < \infty \quad \forall \, x \in C.$$

Summarizing, for any $x \in \mathbb{R}^n$ it holds

$$\mathbb{E}[V(X_t)|X_{t-1} = x] \leq (b_f + b_g\mathbb{E}[\|e_t\|_s])V(x) + b_M\mathbf{1}_C.$$

If $b_f + b_g\mathbb{E}[\|e_t\|_s] < 1$, applying Lemma 15.2.8, Theorem 15.0.1 and Theorem 14.0.1 in Meyn and Tweedie [11], we get that $\{X_t\}$ is a geometrically ergodic Markov chain and, when $s \geq 1$, that the moments of the components of the stationary distribution are finite up to order s. ∎

Remark 7 *A simple extension to the previous result can be obtained by taking, for $s \geq 1$, $|||\cdot|||_s$ or $|||\cdot|||_F$ instead of $|||\cdot|||_{1,s}$. More generally, any norm on $\mathbb{R}^n$ (and the corresponding induced matrix norm) could be considered. Indeed, in the proof the only property of the matrix norms that we use is $\|g(x)e_t\|_s \leq |||g(x)|||_{1,s} \|e_t\|_s$ which holds true for these matrix norms too. Our choice of the matrix norm $|||\cdot|||_{1,s}$ is justified by the possibility to use in assumption iii. absolute moments of order smaller then 1, which weakens the restriction on the noise.*

Example 8 *Let us consider the threshold model of Example 5. We will assume that the i.i.d. random sequence $\{e_t, t \in \mathbb{N}\}$ is distributed according to an Expol(2) law (see [6]) with density $p(x, y) \propto \exp(-(x^2 - 1)^2 - (y^2 - 1)^2)$. This density presents four global maxima and one local minimum and is straightforward to simulate by standard methods. In order to apply the results in Proposition 6, it is easy to check that Assumption ii. hold true, when $s = 1$, for*

$$b_f = \max\{|b_{11}| + |b_{21}|, |b_{12}| + |b_{22}|\}$$

and

$$b_g = \max\{|d_{11}| + |d_{21}|, |d_{31}|, |d_{32}|, |d_{12}| + |d_{22}|\}$$

Since $\mathbb{E}[\|e_1\|_1] \sim 1.66$, it easy to see that the previous model with

$$f(x, y) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$g(x, y) = \begin{pmatrix} 0.1x & -0.15y \\ -0.15x & 0.1y \end{pmatrix} \mathbf{1}_{\mathbb{R}^2 \setminus C} + \begin{pmatrix} 0.2x & 0 \\ -0.25y & 0 \end{pmatrix} \mathbf{1}_C + \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

satisfies the assumptions of Proposition 6, with $b_f + b_g\ \mathbb{E}[\|e_1\|_1] \sim 0.981$. On the other hand, if we modify the coefficients for the conditional mean, for example taking $b_{11} = b_{22} = 1, b_{12} = b_{21} = 0$, the previous assumptions are no more satisfied and the distribution of the solution process does not converge to any distribution, as can be seen by simulation. The same result is obtained if we modify the coefficients for the conditional variance, for example taking

$d_{11} = d_{12} = d_{21} = d_{22} = 0.4$. *Nonetheless simulation of the limit distribution for other set of values shows that even if the sufficient condition is (slightly) violated, the model still appear ergodic, but as pointed out before, Assumption iii. is strong.*

5 Multivariate BEKK-ARCH(1) models with nonlinear autoregressive terms

In this final section we will consider a locally degenerate multivariate BEKK-ARCH(1) model, with a nonlinear autoregressive term. This model belongs to the multivariate BEKK-GARCH class, first proposed by Engle and Kroner in [4], which is particularly useful in multivariate financial time-series, since allow to model both the variances and the covariances. Contrary to all previous works, we will ask that the matrix valued coefficient will be just positive semidefinite and we will be able to derive simple sufficient conditions for the regularity and geometric ergodicity of the solution process.

Let us consider a process $\{X_t, t \geq 0\}$, solution to the following difference equation:

$$X_t = f(X_{t-1}) + (B + (AX_{t-1})(AX_{t-1})^T)^{1/2} e_t, \quad t \geq 1 \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$, B is a positive semidefinite $n \times n$ real matrix, A is a $n \times n$ real matrix and $\{e_t, t \in \mathbb{N}\}$ is a sequence of independent, identically distributed n-dimensional random vectors. Since $(Ax)(Ax)^T$ is, for any $x \in \mathbb{R}^n$, a positive semidefinite $n \times n$ real matrix, $B + (Ax)(Ax)^T$ is itself a positive semidefinite matrix and there exists a unique, positive semidefinite square root (see [8], Chapter 7). The process X_t, solution to (4) is called in the time-series literature a BEKK-ARCH(1) model and its ergodicity has been considered just for the regular case, i.e. assuming that the matrix $B + (Ax)(Ax)^T$ is positive definite and its smaller eigenvalue is uniformly bounded away from zero (see Saikkonon [12]).

To determine a set of sufficient conditions to be the solution process a λ_n-irreducible, aperiodic, T-chain, we will consider for simplicity the case $n = 2$. Let us assume that the matrix B will be non zero and positive semidefinite, but not positive definite. In this case the matrix

$(B+(Ax)(Ax)^T)^{1/2}$ will be positive semidefinite, but could be not positive definite. Indeed, the determinant of the matrix valued function $g(x) = B + (Ax)(Ax)^T$ is equal to

$$\det(g(x)) = b_{11}(a_{21}x_1 + a_{22}x_2)^2 + b_{22}(a_{11}x_1 + a_{12}x_2)^2 - 2b_{12}(a_{21}x_1 + a_{22}x_2)(a_{11}x_1 + a_{12}x_2)$$

and a simple computation show that this determinant is zero for any $(x_1, x_2) \in \mathbb{R}^2$ if both $a_{11}b_{22}^{1/2} - a_{21}b_{11}^{1/2}$ and $a_{12}b_{22}^{1/2} - a_{22}b_{11}^{1/2}$ are zero, while otherwise it is zero just on a straight line $L = \{(x_1, x_2) \in \mathbb{R}^2, \alpha x_1 = \beta x_2\}$, for suitable constants α and β. With the notation of Section 2, we get that in the latter case the set of regular points of g, Θ, coincides with L^c. Moreover, the function $x \to (B + (Ax)(Ax)^T)^{1/2}$ is continuous if $B + (Ax)(Ax)^T$ is positive definite.
A set of sufficient conditions for the assumptions (H.1) and (H.2) will be as follows:

(B.1) At least one between $a_{11}b_{22}^{1/2} - a_{21}b_{11}^{1/2}$ and $a_{12}b_{22}^{1/2} - a_{22}b_{11}^{1/2}$ is different of zero and for any $x \in L = \{x \in \mathbb{R}^2 : \det(g(x)) = 0\}$, there exist $u, v \in \mathbb{R}$ such that $y = f(x) + g(x)u \in \mathcal{C}_f \cap L^c$ and $f(y) + g(y)v \in (\mathcal{C}_f \cap L^c)^o$, with $\mathcal{C}_f^o \neq \emptyset$.

Remark 9 *Clearly, for a specific choice of the autoregressive term f, one can provide better sufficient conditions in order to be the assumptions (H.1) and (H.2) satisfied.*

The next result follows as a simple corollary of previous Proposition 4:

Proposition 10 *Let $n = 2$ and assume that $(B.1)$ and $(H.3)$ are satisfied. Then, the process solution to (4) is a λ_n-irreducible, aperiodic T-chain.*

Let us now consider the geometric ergodicity: in order to apply the results of the previous sections, we will use here the Frobenius matrix norm. In fact we will make use of the fact that $\|A\|_F = \left(\sum_{i,j=1}^n a_{ij}^2\right)^{1/2} = \left(tr(A^T A)\right)^{1/2}$, which gives in our case that

$$\left\| (B + (Ax)(Ax)^T)^{1/2} \right\|_F^2 = tr(B) + tr((Ax)(Ax)^T).$$

A set of sufficient condition for the geometrically ergodicity of the present model follows as a simple modification of previous Proposition 6, whose simple proof we omit.

Proposition 11 *Let* $\{X_t\}$ *be the solution process of (4) and assume that it is a* λ_n*-irreducible, aperiodic, T-chain. If*

i. f *is locally bounded;*

ii. there exist $a_f \geq 0$ *and* $M, b_f > 0$ *such that*

$$\|f(x)\|_2 \leq a_f + b_f \|x\|_2$$

for any $x \in \mathbb{R}^n$ *with* $\|x\|_2 > M$*;*

iii. $\gamma = b_f \ + \ \|A\|_F \ \mathbb{E}\left[\|e_1\|_2\right] < 1$;

then $\{X_t, t \in \mathbb{N}\}$ *is geometrically ergodic.*

Furthermore, the components of the stationary distribution have finite second moments.

Remark 12 *More general conditions for geometric ergodicity are present in the literature of the BEKK-ARCH models (see, for instance, [12]), but a basic ingredient of all the proofs is that the matrix* $B + (Ax)(Ax)^T$ *will be positive definite.*

References

[1] Daren B. H. Cline. Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models. *Adv. in Appl. Probab.*, 39(2):462–491, 2007.

[2] Daren B. H. Cline and Huay-min H. Pu. Verifying irreducibility and continuity of a nonlinear time series. *Statist. Probab. Lett.*, 40(2):139–148, 1998.

[3] Daren B. H. Cline and Huay-min H. Pu. A note on a simple Markov bilinear stochastic process. *Statist. Probab. Lett.*, 56(3):283–288, 2002.

[4] Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150, 1995.

[5] Marco Ferrante, Giovanni Fonseca, and Paolo Vidoni. Geometric ergodicity, regularity of the invariant distribution and inference for a threshold bilinear Markov process. *Statist. Sinica*, 13(2):367–384, 2003.

[6] Marco Ferrante and Federica Giummolé. Finite-dimensional filters for a discrete-time non-linear system with generalized Gaussian white noise. *Stochastics Stochastics Rep.*, 53(3-4):195–211, 1995.

[7] Ernst Hansen and Andres Rahbek. *Stationary and Asymptotics of Multivariate ARCH Time Series with an Application to Robustness of Cointegration Analysis.* Working paper series. Centre for Analytical Finance, Aarhus University, 1998.

[8] Roger A. Horn and Charles R. Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[9] Dennis Kristensen. On stationarity and ergodicity of the bilinear model with applications to GARCH models. *J. Time Series Anal.*, 30(1):125–144, 2009.

[10] Eckhard Liebscher. Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Ser. Anal.*, 26(5):669–689, 2005.

[11] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[12] Pentti Saikkonen. Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity. *Statist. Sinica*, 17(1):221–239, 2007.

[13] H. Tong. A note on a Markov bilinear stochastic process in discrete time. *J. Time Ser. Anal.*, 2(4):279–284, 1981.